

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Appellant:	Bin ZHANG	§	Confirmation No.:	5792
		§		
Serial No.:	10/694,367	§	Group Art Unit:	2168
		§		
Filed:	10/27/2003	§	Examiner:	J. A. Morrison
		§		
For:	Data Mining Method	§	Docket No.:	200310832-1
	And System Using	§		
	Regression Clustering	§		

APPEAL BRIEF

Mail Stop Appeal Brief – Patents

Commissioner for Patents
PO Box 1450
Alexandria, VA 22313-1450

Date: April 9, 2007

Sir:

Appellant hereby submits this Appeal Brief in connection with the above-identified application. A Notice of Appeal was filed via facsimile on February 16, 2007.

TABLE OF CONTENTS

I.	REAL PARTY IN INTEREST	3
II.	RELATED APPEALS AND INTERFERENCES.....	4
III.	STATUS OF THE CLAIMS	5
IV.	STATUS OF THE AMENDMENTS.....	6
V.	SUMMARY OF THE CLAIMED SUBJECT MATTER.....	7
VI.	GROUND OF REJECTION TO BE REVIEWED ON APPEAL.....	9
VII.	ARGUMENT.....	10
	A. Claims 1-17 and 28-30.....	10
	B. Claims 18-23	10
	C. Claims 24-27	10
	D. Conclusion.....	11
VIII.	CLAIMS APPENDIX.....	12
IX.	EVIDENCE APPENDIX	19
X.	RELATED PROCEEDINGS APPENDIX	26

I. REAL PARTY IN INTEREST

The real party in interest is the Hewlett-Packard Development Company (HPDC), a Texas Limited Partnership, having its principal place of business in Houston, Texas. The Assignment from the inventor to HPDC was recorded on October 27, 2003, at Reel/Frame 014652/0977.

Appl. No. 10/694,367
Appeal Brief dated April 9, 2007
Reply to final Office action of October 17, 2006

II. RELATED APPEALS AND INTERFERENCES

Appellant is unaware of any related appeals or interferences.

III. STATUS OF THE CLAIMS

Originally filed claims: 1-30.

Claim cancellations: None.

Added claims: None.

Presently pending claims: 1-30.

Presently appealed claims: 1-30.

Appl. No. 10/694,367
Appeal Brief dated April 9, 2007
Reply to final Office action of October 17, 2006

IV. STATUS OF THE AMENDMENTS

Appellant attempted to amend various claims after the Final Office Action dated October 17, 2006, but the Examiner did not enter the amendments.

V. SUMMARY OF THE CLAIMED SUBJECT MATTER

With the increase in the amount of data being stored in databases, the need to efficiently and accurately analyze data is increasing. Appellant's disclosure, para. [0002]. Appellant's contribution relates to techniques for efficiently mining data from datasets distributed across multiple locations.

According to the invention of claim 1, a processor-based method comprises selecting a set number of functions correlating variable parameters of a dataset. See e.g., Fig. 2, ref. no. 30 and para. [0025]. The method further comprises clustering the dataset by iteratively applying a regression algorithm and a K-Harmonic Means performance function on the set number of functions to determine a pattern in said dataset. See e.g., Fig. 2 and paras. [0025]-[0030].

According to the invention of claim 9, a storage medium comprises program instructions executable by a processor for selecting a set number of functions correlating variable parameters of a dataset. The program instructions also determine distances between datapoints of the dataset and values correlated with the set number of functions, calculate harmonic averages of the distances, regress the set number of functions using datapoint probability and weighting factors associated with the determined distances, repeating the determining and calculating for the regressed set of functions, compute a change in harmonic averages for the set number of functions prior to and subsequent to the regressing, and reiterating the regressing, repeating and computing upon determining the change in harmonic averages is greater than a predetermined value to thereby determine a pattern in the dataset. See e.g., Fig. 2 and paras. [0025]-[0030].

According to the invention of claim 15, a system comprises an input port configured to receive data and a processor configured to regress functions correlating variable parameters of a set of the data, cluster the functions using a K-Harmonic Mean performance function, and repeat the regressing and clustering sequentially to thereby determine a pattern in the dataset. See e.g., Fig. 2 and paras. [0025]-[0030].

According to the invention of claim 18, a system comprises a plurality of data sources and a means for regressively clustering datapoints from the plurality of data sources without transferring data between the plurality of data sources to thereby determine a pattern in data contained in said data sources. See e.g., Fig. 2 and paras. [0025]-[0030].

According to the invention of claim 24, a system comprises a plurality of data sources each having a processor configured to access datapoints within the respective data source and a central station coupled to the plurality of data sources and comprising a processor. The processors of the central station and plurality of data sources are collectively configured to mine the datapoints of the data sources as a whole without transferring all of the datapoints between the data sources and the central station to thereby determine a pattern in datapoints contained in the data sources. See e.g., Fig. 2 and paras. [0025]-[0030].

According to the invention of claim 28, a processor-based method for mining data comprises independently applying a regression clustering algorithm to a plurality of distributed datasets and developing matrices from probability and weighting factors computed from the regression clustering algorithm. The matrices individually represent the distributed datasets without including all datapoints within the datasets. The method further comprises determining global coefficient vectors from a composite of the matrices and multiplying functions correlating similar variable parameters of the distributed datasets by the global coefficient vectors to thereby determine a pattern in the datasets. See e.g., Fig. 2 and paras. [0025]-[0030].

Appl. No. 10/694,367
Appeal Brief dated April 9, 2007
Reply to final Office action of October 17, 2006

VI. GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL

Whether claims 1-30 are anticipated by Zhang et al. ("K-Harmonic Means-Data Clustering Algorithm," hereinafter the "Zhang Reference").

VII. ARGUMENT

A. Claims 1-17 and 28-30

Appellant selects claim 1 as representative of this claim grouping for purpose of the following argument. Claim 1 requires “iteratively applying a regression algorithm and a K-Harmonic Means performance function on the set number of functions to determine a pattern in said dataset.” Claim 1 thus requires “regression.” The Zhang Reference (authored by the present inventor) does not teach regression and, instead, refers to clustering. Clustering and regression are quite distinct. As proof of this point, Appellant submits the attached Table of Contents and Index from a well-known textbook entitled “Applied Regression Analysis” in the Evidence Appendix. Nowhere in the Table of Contents or in the Index of this regression-based textbook does a reference to “clustering” exist. This evidence proves that clustering and regression are distinct concepts.

Additionally, in the Zhang Reference, the clusters are represented by simple geometric centers. Each cluster is a subset of data surrounding a geometric point. In claim 1, however, clusters are represented by “functions” that correlate parameters of the dataset. As such, the claimed “functions” could be lines, curves, planes, hyperplanes, etc., not geometric centers.

Based on the foregoing, Appellant respectfully submits that the rejections of the claims in this first grouping be reversed, and the claims set for issue.

B. Claims 18-23

Appellant selects claim 18 as representative of this grouping. Claim 18 requires “regressively clustering datapoints.” Because the Zhang Reference does not teach regression as explained above, the Examiner erred in rejecting claim 18. Based on the foregoing, Appellant respectfully submits that the rejections of the claims in this grouping be reversed, and the claims set for issue.

C. Claims 24-27

With regard to claim 24, the Examiner’s Final Office Action quoted the claim language and simply pointed to page 1 of the Zhang reference. Independent claim 24 requires a plurality of data sources and a central station. Each of the plurality of data sources and the central station comprise a processor.

Appl. No. 10/694,367
Appeal Brief dated April 9, 2007
Reply to final Office action of October 17, 2006

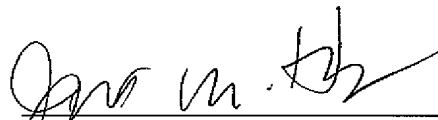
The claim further requires that the processors of the data sources and the central station "are collectively configured to mine the datapoints of the data sources as a whole without transferring all of the datapoints between the data sources and the central station." The Appellant has reviewed page 1 of the Zhang Reference, as well as the rest of the document, and simply does not find a teaching of this combination of limitations.

Based on the foregoing, Appellant respectfully submits that the rejections of the claims in this grouping be reversed, and the grouping set for issue.

D. Conclusion

For the reasons stated above, Appellant respectfully submits that the Examiner erred in rejecting all pending claims. It is believed that no extensions of time or fees are required, beyond those that may otherwise be provided for in documents accompanying this paper. However, in the event that additional extensions of time are necessary to allow consideration of this paper, such extensions are hereby petitioned under 37 C.F.R. § 1.136(a), and any fees required (including fees for net addition of claims) are hereby authorized to be charged to Hewlett-Packard Development Company's Deposit Account No. 08-2025.

Respectfully submitted,



Jonathan M. Harris
PTO Reg. No. 44,144
CONLEY ROSE, P.C.
(713) 238-8000 (Phone)
(713) 238-8008 (Fax)
ATTORNEY FOR APPELLANT

HEWLETT-PACKARD COMPANY
Intellectual Property Administration
Legal Dept., M/S 35
P.O. Box 272400
Fort Collins, CO 80527-2400

VIII. CLAIMS APPENDIX

1. (Previously presented) A processor-based method, comprising:
selecting a set number of functions correlating variable parameters of a dataset; and
clustering the dataset by iteratively applying a regression algorithm and a K-Harmonic Means performance function on the set number of functions to determine a pattern in said dataset.
2. (Original) The processor-based method of claim 1, wherein said clustering comprises:
determining distances between datapoints of the dataset and values correlated with the set number of functions;
regressing the set number of functions using datapoint probability and weighting factors associated with the determined distances;
calculating a difference of harmonic averages for the distances determined prior to and subsequent to said regressing; and
repeating said regressing, determining and calculating upon determining the difference of harmonic averages is greater than a predetermined value.
3. (Original) The processor-based method of claim 2, wherein said determining the distances comprises determining distances from each datapoint of the dataset to values within each function of the set number of functions.
4. (Original) The processor-based method of claim 2, wherein said selecting and said clustering are conducted for a plurality of datasets each from a different data source.
5. (Original) The processor-based method of claim 4, wherein said selecting and said clustering are conducted in parallel for each of the plurality of datasets.

6. (Original) The processor-based method of claim 4, further comprising determining a common coefficient vector to compensate for variations between similar sets of functions within the different data sources.

7. (Original) The processor-based method of claim 6, wherein said determining the common coefficient vector comprises:

- developing matrices from the dataset datapoints and the probability and weighting factors for each of the datasets prior to said reiterating;
- and
- determining the common coefficient vector from a composite of the developed matrices.

8. (Original) The processor-based method of claim 7, further comprising multiplying the similar sets of functions within the different data sources by the common coefficient vector.

9. (Previously presented) A storage medium comprising program instructions executable by a processor for:

- selecting a set number of functions correlating variable parameters of a dataset;
- determining distances between datapoints of the dataset and values correlated with the set number of functions;
- calculating harmonic averages of the distances;
- regressing the set number of functions using datapoint probability and weighting factors associated with the determined distances;
- repeating said determining and calculating for the regressed set of functions;
- computing a change in harmonic averages for the set number of functions prior to and subsequent to said regressing; and

reiterating said regressing, repeating and computing upon determining the change in harmonic averages is greater than a predetermined value to thereby determine a pattern in said dataset.

10. (Original) The storage medium of claim 9, wherein the program instructions are executable using a processor for computing the datapoint probability and weighting factors.

11. (Original) The storage medium of claim 9, wherein the program instructions are executable using a processor for developing matrices from the dataset datapoints and the probability and weighting factors prior to said reiterating.

12. (Original) The storage medium of claim 11, wherein the program instructions are executable using a processor for amassing matrices developed from a plurality of datasets each from a different data source.

13. (Original) The storage medium of claim 11, wherein the program instructions are executable using a processor for determining a common coefficient vector from the composite of matrices.

14. (Original) The method of claim 13, wherein the program instructions are executable using a processor for multiplying similar sets of functions within the different data sources by the common coefficient vector.

15. (Previously presented) A system, comprising:
an input port configured to receive data; and
a processor configured to:
regress functions correlating variable parameters of a set of the data;

cluster the functions using a K-Harmonic Mean performance function; and
repeat said regress and cluster sequentially to thereby determine a pattern in said set of data.

16. (Original) The system of claim 15, wherein the processor is arranged within one of a plurality of data sources each comprising a processor configured to:

regress the functions on a dataset of the respective data source;
cluster the functions using a K-Harmonic Mean performance function; and
repeat said regress and cluster sequentially.

17. (Original) The system of claim 15, further comprising a central station coupled to the plurality of data sources, wherein the central station comprises a processor configured to compute common coefficient vectors which compensate for variations between the regressively clustered functions representing the datasets, and wherein each of the processors of the data sources is configured to alter the functions by the common coefficient vectors.

18. (Previously presented) A system, comprising:
a plurality of data sources; and
a means for regressively clustering datapoints from the plurality of data sources without transferring data between the plurality of data sources to thereby determine a pattern in data contained in said data sources.

19. (Original) The system of claim 18, wherein the means for regressively clustering the datasets comprises a means for applying a regression algorithm and a K-Harmonic Means performance function on the datasets.

20. (Original) The system of claim 18, wherein the means for regressively clustering the datasets comprises a means for applying a regression algorithm and a K-Means performance function on the datasets.

21. (Original) The system of claim 18, wherein the means for regressively clustering the datasets comprises a means for applying a regression algorithm and an Expectation Maximization performance function on the datasets.

22. (Original) The system of claim 18, further comprising a central station communicably coupled to the plurality of data sources, wherein the means is further for:

- collecting dataset information at the central station from the plurality of data sources;
- determining a common coefficient vector from the collected dataset information; and
- altering datasets within the plurality of data sources by the common coefficient vector.

23. (Original) The system of claim 18, wherein the means for regressively clustering the datasets comprises a storage medium with program instructions executable using a processor for:

- selecting a set number of functions correlating variable parameters of a dataset;
- determining distances between datapoints of the dataset and values correlated with the set number of functions;
- regressing the set number of functions using datapoint probability and weighting factors associated with the determined distances;
- calculating a difference of harmonic averages for the distances determined prior to and subsequent to said regressing; and

reiterating said regressing, determining and calculating upon determining the difference of harmonic averages is less than a predetermined value.

24. (Previously presented) A system, comprising:
a plurality of data sources each having a processor configured to access datapoints within the respective data source; and
a central station coupled to the plurality of data sources and comprising a processor, wherein the processors of the central station and plurality of data sources are collectively configured to mine the datapoints of the data sources as a whole without transferring all of the datapoints between the data sources and the central station to thereby determine a pattern in datapoints contained in said data sources.
25. (Original) The system of claim 24, wherein the each of the processors within the plurality of data sources is configured to regressively cluster a dataset within the respective data source.
26. (Original) The system of claim 25, wherein the processor within the central station is configured to:
collect information pertaining to the regressively clustered datasets; and
based upon the collected information, calculate common coefficient vectors which balance variations between functions correlating similar variable parameters of the regressively clustered datasets.
27. (Original) The system of claim 26, wherein the processor within the central station is further configured to:
compute a residual error from the common coefficient vectors;

propagate the common coefficient vectors to the data sources upon computing a residual error value greater than a predetermined value; and

send a message to the data sources to terminate the regression clustering of the datasets upon computing a residual error value less than a predetermined value.

28. (Previously presented) A processor-based method for mining data, comprising:

independently applying a regression clustering algorithm to a plurality of distributed datasets;

developing matrices from probability and weighting factors computed from the regression clustering algorithm, wherein the matrices individually represent the distributed datasets without including all datapoints within the datasets;

determining global coefficient vectors from a composite of the matrices; and

multiplying functions correlating similar variable parameters of the distributed datasets by the global coefficient vectors to thereby determine a pattern in said datasets.

29. (Original) The processor-based method of claim 28, further comprising repeating said independently applying, said developing, said determining and said multiplying.

30. (Original) The processor-based method of claim 28, further comprising calculating a residue error associated with the global coefficients prior to said multiplying.

IX. EVIDENCE APPENDIX

Applied Regression Analysis

THIRD EDITION

Norman R. Draper

Harry Smith




WILEY SERIES IN PROBABILITY AND STATISTICS
REPRINTS AND REVISIONS SECTION

Translated by WALLACE A. STEWART and SAMUEL S. WEISS

Editors: Eric Renshaw, Ralph A. Bradley, Karl A. C. Oakes, Nicholas J. Fisher,
John H. Johnson, J. B. Keane, David G. Kendall, David W. Scott,
Bernard W. Silverman, Adrian R. M. Smith, Josef L. Teugels, Geoffrey S. Watson,
J. Stuart Hunter, Douglas

A complete list of the titles in this series appears at the end of this volume.

Preface	xiii
About the Software	xvii
1 Basic Prerequisite Knowledge	1
1.1 Distributions: Normal, t , and F , 1	
1.2 Confidence Intervals for Means and s^2 -Tests, 4	
1.3 Elements of Matrix Algebra, 6	
2 Fitting a Straight Line by Least Squares	15
2.0 Introduction: The Need for Statistical Analysis, 15	
2.1 Straight-Line Relationship Between Two Variables, 18	
2.2 Linear Regression: Finding a Straight Line by Least Squares, 20	
2.3 The Analysis of Variance, 28	
2.4 Confidence Intervals and Tests for β_0 and β_1 , 34	
2.5 F -Test for Significance of Regression, 39	
2.6 The Correlation Between X and Y , 40	
2.7 Summary of the Straight Line Fit Comparisons, 44	
2.8 Historical Remarks, 45	
Appendix 1A: Student Plots Data, 46	
Exercises are in "Exercises for Chapters 1-3", 56	
3 Checking the Straight Line Fit	47
3.1 Lack of Fit and Pure Error, 47	
3.2 Testing Homogeneity of Pure Error, 50	

This book is prepared on acid-free paper. 

Copyright © 1996 by John Wiley & Sons, Inc. All rights reserved.

For information on this publication, write to: John Wiley & Sons, Inc., 605 Third Avenue, New York, NY 10158-0001.

The paper of this publication meets the requirements of ANSI (American National Standards Institute) permanence guidelines. This publication is printed on acid-free paper.

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data
Dwyer, Nicolas Richard.
Asymptotic regression analysis / N.S. Dwyer. -- 1st ed.
p. cm. -- (John Wiley & Sons, Wiley series in probability and statistics. Texts and monographs in statistics.)
1. Regression analysis. I. Dwyer, N.S. II. Series. III. Title. IV. Texts and monographs in statistics.
QA142.D67 1996
519.5--dc20 96-0001
CIP

ISBN 0-471-58440-8
0-471-58441-6 (pbk.)
0-471-58442-4 (hbk.)

Printed on acid-free paper

vi	CONTENTS	
	Appendix 6A. Orthogonal Columns in the X Matrix, 165	
	Appendix 6B. The Expected Squared Error of Squares, 167	
	Exercises for Chapters 5 and 6, 168	
179	7 Serial Correlation in the Residuals and the Durbin-Watson Test	
	7.1 Serial Correlation in Residuals, 179	
	7.2 The Durbin-Watson Test for a Certain Type of Serial Correlation, 181	
	7.3 Extending Tests to the Time-Sequential Plot of Residuals: Runs Test, 185	
	Exercises for Chapter 7, 186	
205	8 More on Checking Fitted Models	
	8.1 The X Matrix H and the Various Types of Residuals, 205	
	8.2 Added Variable Plot and Partial Residuals, 209	
	8.3 Detection of Influential Observations: Cook's Statistics, 210	
	8.4 Other Statistics Measuring Influence, 214	
	8.5 Reference Checks for Analysis of Residuals, 214	
	Exercises for Chapter 8, 215	
217	9 Multiple Regression: Special Topics	
	9.1 Testing a General Linear Hypothesis, 217	
	9.2 Generalized Least Squares and Weighted Least Squares, 221	
	9.3 An Example of Weighted Least Squares, 224	
	9.4 A Numerical Example of Weighted Least Squares, 226	
	9.5 Restricted Least Squares, 229	
	9.6 Inverse Regression (Multiple Predictor Case), 229	
	9.7 Ridge Regression When All the Variables Are Subject to Error, 231	
	Appendix 9A. Lagrange's Undetermined Multiplier, 231	
	Exercises for Chapter 9, 233	
235	10 Bias in Regression Estimates, and Expected Values of Mean Squares and Sums of Squares	
	10.1 Bias in Regression Estimates, 235	
	10.2 The Effect of Bias on the Least Squares Analysis of Variance, 238	
	10.3 Finding the Expected Values of Mean Squares, 249	
	10.4 Expected Value of Error Sum of Squares, 250	
	Exercises for Chapter 10, 251	
279	11 Fitting Straight Lines: Special Topics	
	11.1 Summary and Preliminaries, 279	
	11.2 Standard Error of $\hat{\beta}$, 281	
	11.3 Inverse Regression (Straight Line Case), 283	
	11.4 Some Practical Design of Experiment Implications of Regression, 285	
	11.5 Straight Line Regression When Both Variables Are Subject to Error, 287	
	Exercises for Chapters 11-13, 289	
315	12 Regression in Matrix Terms: Straight Line Case	
	12.1 Fitting a Straight Line in Matrix Terms, 315	
	12.2 Singularity: What Purpose in Regression to Make $X'X$ Singular? An Example, 323	
	12.3 The Analysis of Variance in Matrix Terms, 327	
	12.4 The Variances and Covariances of $\hat{\beta}_1$ and $\hat{\beta}_2$ from the Matrix Calculation, 330	
	12.5 Variance of $\hat{\beta}$ Using the Matrix Development, 332	
	12.6 Summary of Matrix Approach to Fitting a Straight Line (Nontypical Case), 339	
	12.7 The General Regression Situation, 341	
	Exercises for Chapter 12, 342	
345	13 The General Regression Situation	
	13.1 General Linear Regression, 345	
	13.2 Least Squares Properties, 347	
	13.3 Least Squares Properties When $X \sim N(0, I_p)$, 349	
	13.4 Confidence Intervals Versus Regions, 349	
	13.5 More on Confidence Intervals Versus Regions, 353	
	Appendix 13A. Subspace of Matrix Residuals, 347	
	Exercises for Chapters 12 and 13, 349	
349	14 Extra Sums of Squares and Tests for Several Parameters	
	14.1 Extra Sums of Squares, 349	
	14.2 Tests for Several Parameters, 353	
	14.3 The General Regression Situation, 353	
	14.4 The General Regression Situation, 353	
	14.5 The General Regression Situation, 353	
	14.6 The General Regression Situation, 353	
	14.7 The General Regression Situation, 353	
	14.8 The General Regression Situation, 353	
	14.9 The General Regression Situation, 353	
	14.10 The General Regression Situation, 353	
	14.11 The General Regression Situation, 353	
	14.12 The General Regression Situation, 353	
	14.13 The General Regression Situation, 353	
	14.14 The General Regression Situation, 353	
	14.15 The General Regression Situation, 353	
	14.16 The General Regression Situation, 353	
	14.17 The General Regression Situation, 353	
	14.18 The General Regression Situation, 353	
	14.19 The General Regression Situation, 353	
	14.20 The General Regression Situation, 353	
	14.21 The General Regression Situation, 353	
	14.22 The General Regression Situation, 353	
	14.23 The General Regression Situation, 353	
	14.24 The General Regression Situation, 353	
	14.25 The General Regression Situation, 353	
	14.26 The General Regression Situation, 353	
	14.27 The General Regression Situation, 353	
	14.28 The General Regression Situation, 353	
	14.29 The General Regression Situation, 353	
	14.30 The General Regression Situation, 353	
	14.31 The General Regression Situation, 353	
	14.32 The General Regression Situation, 353	
	14.33 The General Regression Situation, 353	
	14.34 The General Regression Situation, 353	
	14.35 The General Regression Situation, 353	
	14.36 The General Regression Situation, 353	
	14.37 The General Regression Situation, 353	
	14.38 The General Regression Situation, 353	
	14.39 The General Regression Situation, 353	
	14.40 The General Regression Situation, 353	
	14.41 The General Regression Situation, 353	
	14.42 The General Regression Situation, 353	
	14.43 The General Regression Situation, 353	
	14.44 The General Regression Situation, 353	
	14.45 The General Regression Situation, 353	
	14.46 The General Regression Situation, 353	
	14.47 The General Regression Situation, 353	
	14.48 The General Regression Situation, 353	
	14.49 The General Regression Situation, 353	
	14.50 The General Regression Situation, 353	
	14.51 The General Regression Situation, 353	
	14.52 The General Regression Situation, 353	
	14.53 The General Regression Situation, 353	
	14.54 The General Regression Situation, 353	
	14.55 The General Regression Situation, 353	
	14.56 The General Regression Situation, 353	
	14.57 The General Regression Situation, 353	
	14.58 The General Regression Situation, 353	
	14.59 The General Regression Situation, 353	
	14.60 The General Regression Situation, 353	
	14.61 The General Regression Situation, 353	
	14.62 The General Regression Situation, 353	
	14.63 The General Regression Situation, 353	
	14.64 The General Regression Situation, 353	
	14.65 The General Regression Situation, 353	
	14.66 The General Regression Situation, 353	
	14.67 The General Regression Situation, 353	
	14.68 The General Regression Situation, 353	
	14.69 The General Regression Situation, 353	
	14.70 The General Regression Situation, 353	
	14.71 The General Regression Situation, 353	
	14.72 The General Regression Situation, 353	
	14.73 The General Regression Situation, 353	
	14.74 The General Regression Situation, 353	
	14.75 The General Regression Situation, 353	
	14.76 The General Regression Situation, 353	
	14.77 The General Regression Situation, 353	
	14.78 The General Regression Situation, 353	
	14.79 The General Regression Situation, 353	
	14.80 The General Regression Situation, 353	
	14.81 The General Regression Situation, 353	
	14.82 The General Regression Situation, 353	
	14.83 The General Regression Situation, 353	
	14.84 The General Regression Situation, 353	
	14.85 The General Regression Situation, 353	
	14.86 The General Regression Situation, 353	
	14.87 The General Regression Situation, 353	
	14.88 The General Regression Situation, 353	
	14.89 The General Regression Situation, 353	
	14.90 The General Regression Situation, 353	
	14.91 The General Regression Situation, 353	
	14.92 The General Regression Situation, 353	
	14.93 The General Regression Situation, 353	
	14.94 The General Regression Situation, 353	
	14.95 The General Regression Situation, 353	
	14.96 The General Regression Situation, 353	
	14.97 The General Regression Situation, 353	
	14.98 The General Regression Situation, 353	
	14.99 The General Regression Situation, 353	
	15.00 The General Regression Situation, 353	

428	CONTENTS	427
	Appendix 13A How Significant Should My Regression Be?, 247	
	Exercises for Chapter 11, 251	
12	Models Containing Functions of the Predictors, including Polynomial Models 251	
	12.1 More Complicated Model Functions, 252	
	12.2 Worked Examples of Second-Order Surface Fitting for $k = 3$ and $k = 2$ Predictor Variables, 254	
	12.3 Residual Terms in Polynomial Models, 266	
	Exercises for Chapter 12, 272	
13	Transformation of the Response Variable 277	
	13.1 Introduction and Preliminary Remarks, 277	
	13.2 Power Family of Transformations on the Response: Box-Cox Method, 280	
	13.3 A General Method for Estimating λ , 286	
	13.4 Response Transformations: Other Interesting and Sometimes Useful Ones, 289	
	13.5 Other Types of Response Transformations, 290	
	13.6 Response Transformations Chosen to Stabilize Variance, 291	
	Exercises for Chapter 13, 296	
14	"Dummy" Variables 299	
	14.1 Dummy Variables to Separate Blocks of Data with Different Intercepts, Same Model, 299	
	14.2 Interaction Terms Involving Dummy Variables, 307	
	14.3 Dummy Variables for Segmented Models, 311	
	Exercises for Chapter 14, 317	
15	Selecting the "Best" Regression Equation 327	
	15.0 Introduction, 327	
	15.1 All Possible Regressions and "Best Subset" Regression, 329	
	15.2 Stepwise Regression, 335	
	15.3 Backward Elimination, 339	
	15.4 Significance Levels for Selection Procedures, 342	
	15.5 Validation and Summary, 349	
	15.6 Selection Procedures Applied to the Same Data, 349	
	Appendix 15A Model Data, Correlation Matrix, and All 13 Possible Regressions, 348	
	Exercises for Chapter 15, 355	
16	CONTENTS	427
	16.1 Comparing and Testing Regression Lines, 373	
	16.2 Measuring Multicollinearity, 375	
	16.3 Bailey's Suggestion for Detecting Multicollinearity, 396	
	Appendix 16A Transforming X Matrices to Create Orthogonal Columns, 392	
	Exercises for Chapter 16, 395	
17	Ridge Regression 397	
	17.1 Introduction, 397	
	17.2 Basic Form of Ridge Regression, 397	
	17.3 Ridge Regression of the Real Data, 399	
	17.4 In What Circumstances Is Ridge Regression Absolutely the Correct Way to Proceed?, 399	
	17.5 The Phoney Data Viewpoint, 394	
	17.6 Concluding Remarks, 395	
	Appendix 17A Ridge Estimates in Terms of Least Squares Estimates, 396	
	Appendix 17B Mean Square Error Argument, 398	
	Appendix 17C Geometrical Points of Ridge Regression, 397	
	Exercises for Chapter 17, 405	
18	Generalized Linear Models (GLM) 407	
	18.1 Introduction, 407	
	18.2 The Exponential Family of Distributions, 408	
	18.3 Fitting Generalized Linear Models (GLM), 404	
	18.4 Performing the Calculations: An Example, 405	
	18.5 Further Reading, 406	
	Exercises for Chapter 18, 408	
19	Mixture Ingredients as Predictor Variables 409	
	19.1 Mixture Experiments: Experimental Spaces, 409	
	19.2 Models for Mixture Experiments, 422	
	19.3 Mixture Experiments in Randomized Regions, 419	
	19.4 Example 1, 419	
	19.5 Example 2, 419	
	Appendix 19A Transforming k Mixture Variables to $k - 1$ Working Variables, 422	
	Exercises for Chapter 19, 425	
20	The Geometry of Least Squares 427	

20.5 Orthogonalizing in the General Regression Case, 435	26.1 Resampling Procedures for Regression Models, 585
20.6 Ridge Space and Null Space of a Matrix M , 437	26.2 Example: Simple Linear Fit, 586
20.7 The Algebra and Geometry of Pure Error, 439	26.3 Example: Plots for Three Predictors, 588
Appendix 20A. Generalized Inverse M^+ , 441	26.4 Reference Books, 590
Exercises for Chapter 20, 444	Appendix 20A. Sample MINITAB Programs to Bootstrap Residuals for a Specific Example, 589
21 More Geometry of Least Squares	Appendix 20B. Sample MINITAB Programs to Bootstrap Fests for a Specific Example, 591
21.1 The Geometry of a Null Hypothesis: A Single Example, 447	Additional Comments, 591
21.2 General Case: $Hy = Ap + e$: The Projection Algorithm, 449	Exercises for Chapter 26, 591
21.3 Geometric Illustrations, 449	
21.4 The F -Test for H_0 (Geometrically), 450	
21.5 The Geometry of R^2 , 452	
21.6 Change in R^2 for Models Nested Via $A \supset B$ or A Not Involving R_0 , 452	
21.7 Multiple Regression with Two Predictor Variables as a Sequence of Straight Line Regressions, 454	
Exercises for Chapter 21, 459	
22 Orthogonal Polynomials and Summary Data	
22.1 Introduction, 462	
22.2 Orthogonal Polynomials, 462	
22.3 Regression Analysis of Summary Data, 467	
Exercises for Chapter 22, 469	
23 Multiple Regression Applied to Analysis of Variance Problems	
23.1 Introduction, 475	
23.2 The One-Way Classification: Standard Analysis and an Example, 474	
23.3 Regression Treatment of the One-Way Classification Example, 477	
23.4 Regression Treatment of the One-Way Classification Using the Original Model, 481	
23.5 Regression Treatment of the One-Way Classification: Independent Normal Equations, 486	
23.6 The Two-Way Classification with Equal Numbers of Observations in the Cells: An Example, 488	
23.7 Regression Treatment of the Two-Way Classification Example, 488	
23.8 The Two-Way Classification with Equal Numbers of	
23.9 References, 501	
Exercises for Chapter 25, 504	
25 Robust Regression	
25.1 Least Absolute Deviations Regression (L_1 Regression), 507	
25.2 Alternatives, 507	
25.3 Small Example: Example, 515	
25.4 Three Examples, 515	
25.5 Least Median of Squares (LMS) Regression, 517	
25.6 Robust Regression with Residual Residuals (rreg), 517	
25.7 Other Methods, 520	
25.8 Consensus and Outliers, 520	
25.9 References, 501	
Exercises for Chapter 25, 504	
26 Resampling Procedures (Bootstrapping)	
26.1 Resampling Procedures for Regression Models, 585	
26.2 Example: Simple Linear Fit, 586	
26.3 Example: Plots for Three Predictors, 588	
26.4 Reference Books, 590	
Appendix 20A. Sample MINITAB Programs to Bootstrap Residuals for a Specific Example, 589	
Appendix 20B. Sample MINITAB Programs to Bootstrap Fests for a Specific Example, 591	
Additional Comments, 591	
Exercises for Chapter 26, 591	

48	CONTENTS
Tables	684
Normal Distribution, 684	
Percentage Points of the t -Distribution, 686	
Percentage Points of the χ^2 -Distribution, 687	
Percentage Points of the F -Distribution, 688	
Index of Authors Associated with Exercises	685
Index	697

Preface to the Third Edition

The second edition had 18 chapters: 16 chapters, 17a and 17b. On the whole, but not entirely, we have chosen to use smaller chapters, and so distinguish more between different types of material. The table below shows the major relationships between second edition and third edition sections and chapters.

Material dropped consists mainly of second edition Sections 6.8 to 6.13 and 6.13. Sections 7.1 to 7.4 and Chapter 8. New to this edition are Chapters 19 on confidence intervals, 18 on generalized linear models, 19 on regression diagnostics, 20 and 21 on the geometry of least squares, 23 on robust regression, and 26 on resampling procedures. Small revisions have been made even to sections where the text is basically unchanged. Less emphasis has been given to problems, which are now mostly exercises, which are done in the exercises rather than in the problems, which are now mostly exercises. References are mostly given in brief either in text or in the margins, at the end of a section or chapter. Full references are in a bibliography but some references are also given in full in sections or within the text or in exercises, whenever that was felt to be the appropriate thing to do. There is no precise rule for doing this, merely the authors' judgement. Exercises have been grouped as seemed appropriate. They are intended as an exposure to the text and so most exercises have half or partial solutions. There are a very few exceptions, they are marked and one teacher recommends exercises marked. Exercises are revised them to remove "false" questions totally wrong. Sections 26.3 and 26.4 have some duplication with work in Chapter 20, but we decided not to eliminate this because the sections contain some differences and have different exercises. Other smaller differences exist, in general, we find that explanation is a good reason, and so we do not avoid it.

Our hope is that putting this book together in time, it is desirable for students of regression to work through the straight line fit case using a general calculator and then to proceed quickly to studying larger models on the computer. We are aware that

HP PDNO 200310832-1

Appl. No. 10/694,367
Appeal Brief dated April 9, 2007
Reply to final Office action of October 17, 2006

X. RELATED PROCEEDINGS APPENDIX

None.